DOCUMENT RESUME

ED 106 371 TM 004 489

AUTHOR Fleming, Margaret
TITLE Credibility Issues Related to Testing Programs.
PUB DATE [Apr 75]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Credibility; Program Effectiveness; Relevance (Education); *Standardized Tests; Testing; *Testing Problems; *Testing Programs; Tests; Test Selection

ABSTRACT

The implementation of a district-wide testing program requires careful consideration of credibility issues. These issues have been frequently overlooked or treated casually by test developers and users alike. It is well enough to study the reliability and validity issues involved when selecting instrumentation. It is also necessary to identify issues of feasibility and information in ways that garner staff and community support. A review of a district-wide testing program utilized the committee process, staff survey, field testing of all available standardized tests and study of staff development needs in applied measurement. Greater acceptance and more useable information resulted. (Author)

ED106371

CREDIBILITY ISSUES RELATED TO TESTING PROGRAMS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Margaret Fleming
Cleveland Public Schools

TM 004 489

Paper presented at the Annual Meeting of Division H, American Educational Research Association, Washington, D. C., March 30-April 3, 1975

As credibility is commonly defined in dictionary terms--the state of being believed and worthy of belief or trust--the application of this term to selection and implementation processes for standardized testing programs may be readily seen. A growing skepticism related to testing programs is a sign of the times. Although these negative attitudes may reflect common dimensions associated with the decline of public trust in institutions including the school, the deteriorating belief in standardized tests and their role in the educational process has clearly been a major problem faced by those who would utilize tests.

This paper addresses credibility issues generated by procedures for selection and implementation of standardized testing programs for use on a district-wide basis. It presents a practitioner's viewpoint which has been formulated from interactions in the daily operation of a city-wide testing program. As such, it draws upon a belief that how a testing program is installed is of critical importance. This paper predicates that there are practices, particularly in the mechanisms for test selection, that appear to be more effective than others and that testing program administrators need to increase their capability in applying these practices when planning and developing testing programs, if they would maintain credibility for operation of their programs. Finally, there is the basic assumption in this paper that standardized testing programs can be a valuable aid to educators in helping them do a better job for students, provided such tests reflect appropriate technical characteristics and validity for the educational programs in question.

There has been extensive treatment of principles for develop-

ment and selection of standardized tests in educational measurement texts and journals. Criteria for selection and development tests have been definitively presented in the Standards for Educational and Psychological Tests (American Psychological Association 1974). These standards establish guidelines for use and application of instrumentation for test producers and test users alike. A survey of recent representative sources (Anastasi 1968, Bauernfeind 1969, Cronbach 1970, Gronlund 1971, Thorndike and Hagen 1969, Ebel 1972, Stanley and Hopkins 1972, Mehrens and Lehmann 1973) reveals an emphasis on the knowledge base to be applied in test selection. Complete as the literature may be in the matter of technical issues related to tests, relatively few accounts of how tests were selected and by whom, as well as how such programs were implemented have been presented. A recent resource which has systematically delineated a wide range of operational procedures for those administering testing programs is the series, Memo to a Test Director (Ward et al 1973). By and large, however, the major source of information for particulars about selection appear to be annual compilations of testing programs provided by school districts. Two such compilations (Chicago Public Schools 1973 and Milwaukee Public Schools 1974) detail the use of an advisory committee in testing program selection process. Milwaukee's committee included a wide representation of principals, teachers, central office staff, and administrators' and teachers' professional organizations. Ad hoc committees were utilized to extend the representation base to psychologists, counselors, teachers at four different levels and parents. Student opinion was obtained through sessions with school councils. The Chicago Public Schools' report also delineated its use of an Advisory Committee with supporting subcommittees. Their task was to "select tests from among offerings of publishers who

had included the Chicago Public Schools in the standardization processes." In both school systems, committee operations featured a review of available instruments. The Chicago plan also included publisher presentations, while the Milwaukee operation utilized a survey of staff input about current issues in educational measurement to focus on a "needs for data" approach. This latter process led to identification by the staff of major areas in which measurement should be provided by a revised testing program.

One of the rare statements located in the literature about the need to document the processes related to transactions involving the partners in the testing enterprise is this appraisal (Dyer 1973):

> ....Broadly considered, there are four groups of people who are involved in the transactions we call educational testing: the test makers, the test givers, the test takers and the test users. From this view of the enterprise, two observations can be made. First, both within and across the four groups of participants, there is an extraordinary amount of diversity in their understanding of tests and in their attitudes toward testing. And second, as mass testing has spread throughout the schools of the nation, it has become more and more compartmentalized--that is, disjointed--with the result that the interrelationships among the four groups (the makers, the givers, the takers, the users) have become increasingly strained and tenuous. And the consequence of this is that communications among them are becoming more and more like random events.

And Dyer concludes that the immediate task is "to get the solutions out of the literature and translate them into terms that will make them functional in school testing problems."

The framework of this paper represents an attempt to functionalize solutions related to communication about the installation of a city-wide testing program. A recent decision to review the city-wide testing program provided the field setting for implementation of test selection processes, the opportunity to document the operations involved in these

processes and the occasion to provide for study of such operations.

The Cleveland Public Schools periodically institutes a review of its city-wide testing program. The mechanism for such review is a Test Review Committee--which is convened approximately every three to five years. The present Committee was organized to include a total of 22 school, supervisory and administrative personnel. These staff were representative of the school levels and range of curriculum areas included in the city-wide testing program. In accordance with the agreements maintained between the school system and the teachers' union and administrators' and supervisors' council, each professional organization named committee members as their group representatives. The Division of Research and Development was designated as the liaison group to provide supportive services for Committee activities.

The Committee identified four major tasks as essential to its role as a reviewing mechanism. These included:

1. Study of the present rationale, scope and sequence for the city-wide testing program.
2. Review of all major standardized tests in basic skills and scholastic aptitude areas for the grade span from kindergarten through grade 12.
3. Field test of the major test series in a representative sample of Cleveland schools and classes.
4. Consideration of feedback from school staff utilizing the present program and field test program.

The original time frame for Committee activities had been visualized as being a 9-month period. The complex scope of the activities and interest of the Committee, however, resulted in an expansion of this period to a year and a half. A series of meetings scheduled during this period focused on these topics:

1. Study of the history of city-wide testing programs in the school district with particular reference to the rationale behind the most recent program;

2. study of the APA standards for psychological tests;

3. process review of the operations of the present city-wide testing program;

4. study session with national consultant on issues related to test bias;

5. identification of and consensus building for criteria for review of available achievement and aptitude instruments;

6. review and rating of available achievement and aptitude instruments;

7. periodic consultations and hearings, with curriculum specialists, counselors and teachers in special subject areas about content validity of various instruments;

8. preparation of committee report of findings and recommendations.

As a base for operations, the Committee identified a series of questions of concern with which it would deal. As can be seen from Appendix I, these questions ran the gamut of issues related to standardized testing. Next, the activities necessary to provide a decision base for the solutions were scheduled by the Committee. For example, feedback from the school staff utilizing the present program was considered critical to study of these issues. Therefore, a testing program survey was administered to determine staff opinion and recommendations about the present city-wide program. The survey was circulated to a representative sample of teachers, principals, department heads and counselors across the grade levels. A response rate of 61% was obtained.

The survey revealed that the respondents were generally familiar with the city-wide testing program. Generally reaffirmed were the present grade sequences for administration of the achievement and scholastic

aptitude tests. Recommendations from this population for the time of year for testing also supported the present testing schedule. Assessment of the needs for certain staff development activities were also of interest to the Committee in its deliberations. Elementary principals, secondary grade counselors, teachers and department chairmen indicated their needs for staff development activities related to interpretation of results for instruction. Elementary principals, secondary principals and secondary counselors also reported a need for staff development activities to assist them in communicating results to parents and pupils. Such an outcome was anticipated because recent staff development activities have been directed primarily at elementary teachers.

Another on-going activity of the Committee was the review of all the available achievement and scholastic aptitude instruments. Kits of all tests, manuals and/or technical bulletins, provided through the cooperation of the test publishers for the most part, were distributed to Committee members. Each member prepared a rating scale for each test reviewed. An example of the rating scale is presented in Appendix II. The Committee decided upon its own involvement in study of the tests rather than presentations by the publishers. Advice about technical aspects of the instrumentation was provided by staff of the Division of Research and Development.

The committee also viewed the information to be provided from field test of these instruments as most critical to their deliberations. Try-out of the instruments was effected through a field study involving a sample of 5,742 pupils in representative classrooms throughout the district. Sixty-one various reading, mathematics and scholastic aptitude tests were administered across the entire span of grades where various

levels of the instruments would be most appropriately used. ..... These data were related to the results produced by the on-going city-wide program in an attempt to examine correlations and to compare performance data on these tests.

Those teachers administering the instruments prepared ratings of what were considered to be practical features--test format, test length, clarity of administration procedures for pupils, clarity of administration procedures for test administrators, students' use of answer sheets and the like. Teachers made liberal use of the comments section of these rating sheets in supplying their perceptions of the test and testing situation.

The information generated by the teachers conducting the field try-out was collated with the summary of ratings of the test review committee. Table 1 shows a section of the comparison data related to review of seven reading instruments. Together with the performance and correlation data, this information was viewed as critical input for committee deliberations and documentation for its decisions about test selection.

Insert Table 1 here

Information from study of scholastic aptitude instruments by a panel of counselors and a review of specialized curriculum areas of mathematics and science by a subcommittee of curriculum supervisors and representative teachers and department chairmen provided additional feasibility and content validity input to the Committee.

The final task of the Committee was the preparation of the report which is presently in press. The use of standardized testing program as a viable process in the evaluation process of the school program was reaffirmed by the Committee. The major purpose of the testing program

was viewed as the improvement of instruction with the information generated by the testing program being utilized for these purposes:

1. to describe specific learning difficulties of school and class groups;

2. to assess performance levels of groups and individuals against internal and external standards;

3. to provide objective data for use with other information in educational and vocational decision-making.

The focus and hope of the testing program in the Committee's opinion was to contribute not just scores to be duly recorded in school records, but data for decisions related to the instructional program. Quality control in such decision-making was another committee concern. For this reason, the Committee has recommended a diagnostic framework for the use of results. While the classroom teacher was viewed as the prime consumer for test results, it was the Committee's intention that test scores should be reported to all who should know, along with appropriate interpretation of what the scores mean. Such audiences, in the Committee's view, include administrators, counselors, teachers, pupils and parents.

The Committee report provides a primary source document about intents and purposes for the utilization of standardized tests in the school district. In detailing the desired components of operating policies for the testing program, components of the report address a range of major topics, for example:

- rationale for the city-wide program;
- maintenance of standards promulgated by APA, AERA and NCME;
- provisions for appropriate dissemination of the intended use of the program;

- full utilization of computer-generated materials to aid the dissemination and information processes;
- expansion of interpretation services via a range of media to better inform the educational partners;
- continuation of studies of test validity for Cleveland pupils and programming;
- establishment of local policies for dissemination of results to the public.

The recent activities of the Test Review Committee have many implications for the future of the city-wide testing program in Cleveland schools. The Committee has been the mechanism by which processes utilized for decisions about the testing program operation and selection has been documented. Such documentation was considered to be an essential foundation for program credibility.

Committee recommendations have identified certain critical needs for a viable testing program:

- the necessity for resources to support on-going and systematic staff development activities related to test interpretation;
- the desirability for interim reviews of testing program operations through the committee process;
- priorities for policies for systematic dissemination of test results to parents and the public;
- priorities for appropriate communication to pupils and parents and development of computer-generated individualized materials for this purpose;
- the necessity for systematic introduction of the revised testing program through orientation of staff and students.

Future implementation of these recommendations will further the development of the credibility base for the program. In addition, as the new program is implemented, an appropriate process evaluation design will be mounted to assess the degree to which the partners in

the testing process--test givers and test takers, know and feel better about the testing program. It is anticipated that additional credibility issues in matters of dissemination of results to parents and the media should next be considered. Larsen (1974) recently presented a series of useful suggestions for communicating information to professionals, parents, students and the public. Hopefully such processes will further serve to dissipate those "random communication events," which Dyer decribes, in the matters related to testing.

TABLE I

READING

SUMMARY OF RATINGS: TEST REVIEW COMMITTEE
RATINGS OF "INADEQUATE"

Test Content	Test A %	Test B %	Test C %	Test D %	Test E %	Test F %	Test G %
Adequacy of Content Categories	40	--	--	06	--	--	33
Correspondence Between Test Content And Instructional Content	60	14	--	--	--	09	33
Reading Level	40	--	--	--	--	09	50
Practical Features							
Test Format Appearance	--	--	--	06	23	45	--
Test Length	38	25	13	06	08	10	33
Clarity of Administration Procedures for Pupil	33	18	--	--	23	20	--
Clarity of Administration Procedures for Test Administrator	33	--	--	--	08	18	--

SUMMARY OF RATINGS: FIELD TEST STAFF
RATINGS OF "PROBLEMS"

Practical Features	ITBS* %	Test A %	Test B %	Test C %	Test D %	Test E %	Test F %	Test G %
Test Format/ Appearance	67	--	17	11	--	8	25	--
Test Length	48	20	4	--	--	50	25	17
Clarity of Administration Procedures for Pupil	52	--	9	11	--	21	12	--
Clarity of Administration Procedures for Test Administrator	4	--	4	--	17	8	--	33
Students' Use of Answer Sheets	64	--	4	--	17	8	--	33

* ITBS - Designs for Learning Project

APPENDIX I

SUGGESTED QUESTIONS OF CONCERN

1. In what ways, if any, can standardized testing programs contribute to the educational program?

2. Are so-called "scholastic aptitude tests" of use to the district program?

3. What elements, if any, of the present standardized testing program appear to be worth continuing?

4. What guidance should be considered from the proposed Testing Standards prepared by APA, AERA and NCME?

5. What achievement areas, if any, should be included in a standardized testing program?

6. If used, when should standardized tests be scheduled in terms of grade sequence and the school calendar?

7. What norms comparison plan should be utilized for standardized testing programs? (national, large city, local?)

8. What precautions against bias in assessment of district pupils should be insured?

9. Are there test series that are appropriate for the Cleveland schools?

10. Should a longitudinal or cross sectional plan be used in relation to the test program?

11. What staff development efforts appear critical? How should these procedures be implemented?

12. What feedback should be given? Who should receive information about results?

13. In what form should feedback be provided?

14. What interpretation services are required?

15. What use of test results should be encouraged?

16. After all is said and done, should criterion-referenced testing be included in the district program?

APPENDIX II

RATING SCALE FOR STANDARDIZED TESTS

Test Name_____ Level___ Form___ Section_____

	Excellent	Adequate	Inadequate	No Data
Test Content				
Rationale For Structure of Test	___	___	___	___
Adequacy of Content Categories	___	___	___	___
Correspondence Between Test Content and Instructional Content	___	___	___	___
Reading Level	___	___	___	___
Norms				
Appropriateness of Norming Sample	___	___	___	___
Multiple-Norm-Group Data:	___No	___Yes	Groups: _____	
Practical Features				
Test Format/Appearance	___	___	___	___
Test Length: _____Min.	___	___	___	___
Clarity of Administration Procedures for Pupil	___	___	___	___
Clarity of Adminstration Procedures for Test Administrator	___	___	___	___
Clarity of Scoring Procedures	___	___	___	___

Equivalent Forms: No. Provided _____

Converted Scores: Types Provided (Check)

___ Grade-Equivalent ___ Standard Score

___ Percentile ___ Stanine

Score Adjustment For Time of Testing: ___No ___Yes

Response Modes:

___ Machine-Scorable Booklet

___ Separate Answer Sheet

Types Available: _____

No. of Time-Points: _____

BIBLIOGRAPHY

American Psychological Association. Standards for Educational and Psychological Tests. Washington, D. C., 1974.

Anastasi, A. Psychological Testing. 3rd ed. New York: Crowell Collier and MacMillan, Inc., 1968.

Bauernfeind, R. H. Building a School Testing Program. Boston: Houghton Mifflin Company, 1969.

Chicago Public Schools. Report on the City-Wide Testing Program 1972-73. Chicago: Board of Education, April 1974.

Cronback, L. J. Essentials of Psychological Testing. 3rd ed. New York: Harper and Row, Publishers, 1970.

Dyer, H. E. "Recycling the Problems in Testing." Proceedings of the 1972 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1973.

Ebel, R. L. Essentials of Educational Measurement. Englewood Cliffs: Prentice-Hall, Inc., 1972.

Gronlund, N. E. Measurement and Evaluation in Teaching. 2nd ed. New York: MacMillan, 1971.

Larsen, E. P. "Opening Institutional Ledger Books--A Challenge to Educational Leadership." TM Report No. 28, ERIC Clearinghouse on Tests, Measurement & Evaluation. Princeton: Educational Testing Service, 1974.

Mehrens, W. A. and Lehmann, I. L. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, Inc., 1973.

Milwaukee Public Schools. Revision of City-Wide Testing Program 1973-1974. Milwaukee: Division of Planning and Long-Range Development, 1974.

Stanley, J. C. and Hopkins, K. D. Educational and Psychological Measurement and Evaluation. Englewood Cliffs: Prentice-Hall, Inc., 1972.

Thorndike, R. L. and Hagen, E. Measurement and Evaluation in Psychology and Evaluation. 3rd ed. New York: John Wiley and Sons, Inc., 1969.

Ward, A., editor, "Memo to a Test Director." Measurement News 16 (1973) 1-3.